Depth Map Down-Sampling and Coding Based on Synthesized View Distortion

Chao Yao, Jimin Xiao, Tammam Tillo, *Senior Member, IEEE*, Yao Zhao, *Senior Member, IEEE*, Chunyu Lin, and Huihui Bai

***Abstract*—In this paper, we propose a depth map down-sampling and coding scheme that minimizes the view synthesis distortion. Moreover, a solution for the optimal depth map down-sampling problem that minimizes the depth-caused distortion in the virtual view by exploiting the depth map and the associated texture information along with the up-sampling method to be used in the decoder side is derived. Furthermore, to enhance compression performance, the synthesized view distortion, which is evaluated by emulating the interpolation and the virtual view synthesis process, is used in the optimization objective function for coding mode selection in the video encoder. Experimental results show that both the proposed depth map down-sampling and encoding methods lead to good performance, and the average bit rate reduction is 2.62% compared with 3D-AVC.**

***Index Terms*—Depth coding, depth down-/ up-sampling, optimal depth down-sampling, three-dimensional (3D), view synthesis.**

I. Introduction

In THE past decade, 3D video has attracted attention from both industry and academia. Many applications, such as 3D TV and free-view TV, are entering the market. In general, 3D video can be represented by Multi-View plus Depth (MVD) format. MVD format consists of multi-view texture videos and associated-per-pixel depth maps [1]. The depth map represents the distance from the camera to the objects in the scene. It has been widely applied to generate the virtual views with the Depth Image-Based Rendering method [2].

Depth maps are, usually, represented by 8-bit gray-scale value. However, different from the natural videos or images, depth maps generally have more homogeneous regions compared with the corresponding texture videos or images [3]. Recent studies [4]–[6] have also shown that depth distortion has less perceptual impact than color distortion on the final quality of the synthesized views. Based on this property, resolution-reduction techniques of the depth map, represent an efficient approach for compression [7]. So for example, 3D-AVC [8] which is an H.264/MVC-based test model for 3D video coding standardization, encodes depth map data at a reduced resolution as the default setting, where depth maps are firstly down-sampled prior to encoding and in the decoder side bilinear filter is used to up-sample the decoded depth maps to the full resolution. However, given that depth down-sampling would also introduce some distortion, efficient depth map resolution-reduction and pre/post-processing methods are necessary to guarantee high synthesized view quality. Classical down/up-sampling methods (such as low-pass filters and linear interpolation filters) have been used for depth coding [9]. However, in these methods, the lost high frequency information produces jagged boundaries and annoying artifacts around the edges of foreground objects. In order to maintain edge sharpness, in [10] a depth reconstruction filter was proposed to recover object boundaries, whereas Liu *et al.* proposed a joint trilateral filter [11] to post-process the reconstructed depth images. Deng *et al.* in [12], proposed an edge-preserving interpolation method for up-sampling-based depth coding, which uses edge similarity between the depth map and its corresponding texture image to suppress artifacts due to down-sampling and coding. Similarly, to improve the synthesized view quality some approaches exploit the texture information, so for example, in [13], a color-based depth map up-sampling approach was proposed, which takes into account the high-resolution texture information. Liu *et al.* [14] proposed a cross-view down/up-sampling method, which exploits cross-view information to assist the up-sampling at the decoder.

In additional, to apply the resolution-reduction depth coding approach in 3D-AVC, Aflaki *et al.* proposed a nonlinear depth resampling method in [15], [16]. To estimate the sampling value, the nonlinear resampling method firstly divides the entire image into some non-overlap blocks. Then, based on the closeness-favored averaging algorithm, the blocks are classified as foreground/background blocks. Finally, for each block, by considering the property of the block and the distribution of foreground/background pixels in the block, one sampled value is estimated to represent the correspondence block in the down-sampled image. By doing this, the method can preserve sharp boundaries of foreground objects. Although these methods achieve relatively good performance, they mainly focus on recovering the lost high frequency information (edges of objects in depth maps), and they do not provide an optimal

Manuscript received August 30, 2015; revised November 9, 2015 and June 13, 2016; accepted July 14, 2016. Date of publication July 27, 2016; date of current version September 15, 2016. This work was supported in part by the 973 Program under Grant 2012CB316401, in part by the National Natural Science Foundation of China under Grant 61210006, Grant 61501379, Grant 61502506, and Grant 61402034, and in part by the Jiangsu Science and Technology Programme under Grant K20150375. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Chia-Wen Lin.

C. Yao, C. Lin, and H. Bai are with the Institute of Information Science and the Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing Jiaotong University, Beijing 100044, China (e-mail: yaochao1986@gmail.com; hhbai@bjtu.edu.cn; cylin@bjtu.edu.cn).

J. Xiao and T. Tillo are with the Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China (e-mail: tammam.tillo@xjtlu.edu.cn; Jimin.Xiao@xjtlu.edu.cn).

Y. Zhao is with the Institute of Information Science, the State Key Laboratory of Rail Traffic Control and Safety, and the Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing Jiaotong University, Beijing 100044, China (e-mail: yzhao@bjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2016.2594145

solution of the depth map down/up-sampling operation. Thus it is not easy to predict their performance. In order to minimize the Mean Squared Error between the input image and the interpolated image, in [17] Zhang *et al.* introduced a dependent down-sampling algorithm for a given interpolation method. However, since depth distortion doesn't affect the synthesis distortion in a linear fashion, applying this approach directly for depth map down-sampling would not be optimal. In [18], [19] a new distortion metric was introduced to measure the impact of depth map error on the synthesized view quality; later Oh *et al.*, in [20], further improved the accuracy of this distortion metric by involving the texture information of the video, which has been accepted as view synthesis distortion (VSD) in 3DV standardization, and it is used to improve the Rate-Distortion performance during depth encoding [21], [22]. However, this metric does not analyze the effect of depth map down sampling.

In this paper, an optimal depth down-sampling and coding scheme is proposed, where we try to minimize the VSD caused by errors of depth map down/up-sampling and encoding process. In the depth down-sampling process, the depth interpolation method which will be used in the decoder side and the corresponding texture information are used to evaluate the VSD, so as to optimally select the down-sampled data that minimizes the VSD. Furthermore, in the depth reduced-resolution coding stage, a modified R-D cost function is used for coding mode decision in 3D-AVC, where the used distortion function includes the VSD caused by both the depth up/down-sampling and encoding process.

The rest of this paper is organized as follows. Section II discusses the proposed depth down-sampling and coding algorithm, where the conventional resolution-reduction coding scheme is presented in Section II-A, the proposed optimal depth down-sampling is described in Section II-B. In Section II-C the proposed down-sampled depth coding scheme is presented. The experimental results are reported in Section III. Finally, this paper is concluded in Section IV.

II. Methodology

A. Problem Formulation

In [23] it has been shown that depth down-sampling operator prior to compression and up-sampling after decoding can improve the coding performance. In 3D-AVC, prior to encoding, original depth maps are firstly down-sampled to low resolution, where the horizontal and vertical ratios are 2. In the decoder side, the low resolution depth maps are constructed from the encoded video bitstream and then up-sampled to the full resolution by using bilinear interpolation method. To evaluate the overall distortion, the VSD proposed in [20] has been adopted in 3D-AVC and 3D-HEVC. It can be denoted as

$$\begin{aligned} VSD &= \|\mathbf{V} - \hat{\mathbf{V}}\|^2 \\ &= \|f_w(\mathbf{C}, \mathbf{D}) - f_w(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{D}}})\|^2 \end{aligned} \tag{1}$$

where $\mathbf{V}$ and $\hat{\mathbf{V}}$ represent the virtual images generated by a pre-defined 3D warping function, f_w, using the original and reconstructed texture image $\mathbf{C}$ and $\hat{\mathbf{C}}$. In addition, the co-located full-resolution original depth map $\mathbf{D}$ and the reconstructed depth map $\hat{\tilde{\mathbf{D}}}$ are also used. Here, it should be noted that in our resolution-reduction coding system $\hat{\mathbf{C}}$ is the decoded texture signals and $\hat{\tilde{\mathbf{D}}}$ is the reconstructed depth signals after being decoded including down/up-sampling stages. Correspondingly, with considering the depth down/up-sampling distortion, the VSD function can be represented as

$$VSD = \|f_w(\mathbf{C}, \mathbf{D}) - f_w(\mathbf{C}, \tilde{\mathbf{D}}) + f_w(\mathbf{C}, \tilde{\mathbf{D}}) - f_w(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{D}}})\|^2 \tag{2}$$

where $\tilde{\mathbf{D}}$ represents the reconstructed depth map after down/up-sampling. Here, the overall distortion includes: depth down/up-sampling distortion $f_w(\mathbf{C}, \mathbf{D}) - f_w(\mathbf{C}, \tilde{\mathbf{D}})$ and coding-caused distortion $f_w(\mathbf{C}, \tilde{\mathbf{D}}) - f_w(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{D}}})$. The two types of distortions can be viewed as uncorrelated, due to that they are caused by different reasons.

We designed an experiment to verify that the two types of distortions are uncorrelated. Firstly, the original depth map $\mathbf{D}$ is down-sampled and up-sampled to generate the low resolution version $\tilde{\mathbf{d}}$ and the up-sampled version $\tilde{\mathbf{D}}$.[1] Then the error $f_w(\mathbf{C}, \mathbf{D}) - f_w(\mathbf{C}, \tilde{\mathbf{D}})$ on each pixel can be calculated, and this error is caused by depth down/up-sampling; Secondly, the original texture image $\mathbf{C}$ and depth map $\tilde{\mathbf{d}}$ are encoded and decoded by using 3D-AVC, then the decoded texture $\hat{\mathbf{C}}$ and the decoded-up-sampled depth map $\hat{\tilde{\mathbf{D}}}$ are used to count the error $f_w(\mathbf{C}, \tilde{\mathbf{D}}) - f_w(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{D}}})$ on each pixel. Note that $f_w(\mathbf{C}, \tilde{\mathbf{D}})$ is used as the reference to count the error caused by coding. Finally, the correlation coefficient between $f_w(\mathbf{C}, \mathbf{D}) - f_w(\mathbf{C}, \tilde{\mathbf{D}})$ and $f_w(\mathbf{C}, \tilde{\mathbf{D}}) - f_w(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{D}}})$ on each pixel is computed. By averaging value of all pixels in each frame, depth-down/up-sampling-caused distortion and coding-caused distortion can be proved as uncorrelated. In this experiment, we take 30 frames of sequences "*Newspaper*", "*Kendo*", "*Poznan_Streetand*" and "*Undo_Dancer*" as examples, and the correlation coefficient for each frame is shown in Fig. 1. It shows that the correlation

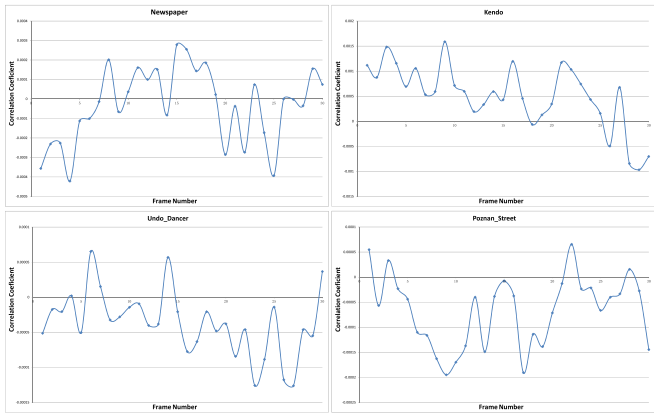

Fig. 1. Correlation coefficients between $f_w(\mathbf{C}, \mathbf{D}) - f_w(\mathbf{C}, \tilde{\mathbf{D}})$ and $f_w(\mathbf{C}, \tilde{\mathbf{D}}) - f_w(\hat{\mathbf{C}}, \hat{\tilde{\mathbf{D}}})$ for each frame.

[1] Here, we use bilinear method for sampling.

coefficient on each frame is very close to 0, for example the correlation coefficients of sequence "*Undo_Dancer*" are in the range of $(-0.00015, 0.00015)$. Therefore, the correlation coefficients between depth-down/up-sampling-caused distortion and coding-caused distortion are approximately 0.

Therefore, it means that $(f_w(\mathbf{C},\mathbf{D}) - f_w(\mathbf{C},\tilde{\mathbf{D}}))(f_w(\mathbf{C},\tilde{\mathbf{D}}) - f_w(\hat{\mathbf{C}},\hat{\tilde{\mathbf{D}}}))$ is 0, so (2) can be written as

$$\begin{aligned} VSD = & \|f_w(\mathbf{C},\mathbf{D}) - f_w(\mathbf{C},\tilde{\mathbf{D}}) + f_w(\mathbf{C},\tilde{\mathbf{D}}) - f_w(\hat{\mathbf{C}},\hat{\tilde{\mathbf{D}}})\|^2 \\ = & \|f_w(\mathbf{C},\mathbf{D}) - f_w(\mathbf{C},\tilde{\mathbf{D}})\|^2 \\ & + \|f_w(\mathbf{C},\tilde{\mathbf{D}}) - f_w(\hat{\mathbf{C}},\hat{\tilde{\mathbf{D}}})\|^2. \end{aligned} \tag{3}$$

At last, the optimization problem, i.e., to improve the overall performance in the depth-resolution-reduction coding system, can be formulated as two sub-problems: depth down/up-sampling optimization $\|f_w(\mathbf{C},\mathbf{D}) - f_w(\mathbf{C},\tilde{\mathbf{D}})\|^2$ and depth coding optimization $\|f_w(\mathbf{C},\tilde{\mathbf{D}}) - f_w(\hat{\mathbf{C}},\hat{\tilde{\mathbf{D}}})\|^2$.

B. Depth Down-Sampling Optimization

In [17] an optimal natural image down/up-sampling method is proposed. This uses a paradigm that an optimized down/up-sampling process should guarantee that the up-sampled version is as close as possible to the original non down-sampled version, it can be denoted as

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \|\tilde{\mathbf{Y}} - \mathbf{Y}\|^2 \tag{4}$$

where $\mathbf{Y}$ denotes the full-resolution original natural image with size $M \times N$, and $\mathbf{X}$ the down-sampled image with size $M/2 \times N/2$, whereas, $\tilde{\mathbf{Y}}$ denotes the interpolated image, i.e., the up-sampled image to size $M \times N$. The interpolation process of the low-resolution image[2] can be expressed mathematically as

$$\tilde{\mathbf{Y}}_c = \mathbf{H}\,\mathbf{X}_c \tag{5}$$

where $\mathbf{H}$ is the interpolation matrix with size $(MN) \times (MN/4)$, and its coefficients are generated for each specific interpolation method. The subscript c stands for the *column vector* version of a reshaped matrix, so for example $\mathbf{X}_c$ is a column vector of size $(MN)/4 \times 1$ containing the reshaped down-sampled image $\mathbf{X}$, and $\tilde{\mathbf{Y}}_c$ is the column vector of $\tilde{\mathbf{Y}}$. Similarly, in the following description, $\tilde{D}_c$ is the column vector of the high resolution depth map $\mathbf{D}$, $\mathbf{d_c}$ is column vector of down-sampled depth map d.

In our proposed framework, based on (3), the target of the depth down-sampling optimization is to minimize $\|f_w(\mathbf{C},\mathbf{D}) - f_w(\mathbf{C},\tilde{\mathbf{D}})\|^2$. Since depth distortion does not affect the synthesis distortion in a linear fashion, the VSD is used instead of the depth distortion. Hence, in this work, we optimize the depth map down-sampling process by modifying (4) to account for the distortion in the synthesized view as follows:

$$\begin{aligned} \mathbf{d}^* &= \arg\min_{\mathbf{d}} \|\mathbf{V} - \tilde{\mathbf{V}}\|^2 \\ &= \arg\min_{\mathbf{d}} \|f_w(\mathbf{C},\mathbf{D_c}) - f_w(\mathbf{C},\tilde{\mathbf{D}}_\mathbf{c})\|^2 \\ &\quad \text{where } \tilde{\mathbf{D}}_\mathbf{c} = \mathbf{H}\,\mathbf{d_c} \end{aligned} \tag{6}$$

[2]Image in this context could refer to texture image or depth map image.

where $\mathbf{d}^*$ denotes the optimal down-sampled depth version. In the following, let $C_{x,y}$, $D_{x,y}$ and $\tilde{D}_{x,y}$ represent the values of $\mathbf{C}$, $\mathbf{D}$ and $\tilde{\mathbf{D}}$ at position (x,y), respectively. For simplicity hereinafter the term $\|f_w(\mathbf{C},\mathbf{D}) - f_w(\mathbf{C},\tilde{\mathbf{D}})\|^2$ is denoted by E_d. Following [18], the 1-D parallel camera setting configuration is set as default, thereby, (6) can be approximately written as

$$\begin{aligned} E_d &= \sum_{\forall(x,y)} \|f_w(\mathbf{C},\mathbf{D}) - f_w(\mathbf{C},\tilde{\mathbf{D}})\|^2 \\ &\approx \sum_{\forall(x,y)} \|\mathbf{C}_{x,y} - \mathbf{C}_{x-\Delta p(x,y),y}\|^2 \end{aligned} \tag{7}$$

where Δp denotes the translational rendering position, which has been proven to be proportional to the depth map error

$$\Delta p(x,y) = \alpha \cdot (D_{x,y} - \tilde{D}_{x,y}) \tag{8}$$

where α is a proportional coefficient determined by the following equation:

$$\alpha = \frac{f \cdot L}{255} \cdot \left(\frac{1}{Z_{\text{near}}} - \frac{1}{Z_{\text{far}}}\right) \tag{9}$$

where f is the focal length, L is the baseline between the current view and the rendered view, Z_{near} and Z_{far} are the values of the nearest and farthest depth of the scene, respectively.

The VSD, described in (7), could be further simplified according to [20] as

$$E_d \approx \sum_{\forall(x,y)} \left[|\Delta p(x,y)| \frac{|C_{x,y} - C_{x-1,y}| + |C_{x,y} - C_{x+1,y}|}{2} \right]^2. \tag{10}$$

In this equation the value of $C_{x,y}$ is set to zero when x exceeds the size of the image. It is possible to note from this equation the effect of depth error and texture complexity on the synthesized view distortion. The distortion calculation requires $\Delta p(x,y)$ which could be obtained by rewriting (8) as

$$\Delta \mathbf{p}_c = \alpha \cdot (\mathbf{D}_c - \mathbf{H}\,\mathbf{d_c}) \tag{11}$$

where $\mathbf{H}\,\mathbf{d_c}$ is the interpolated version of the depth map, which has been evaluated according to (5).

Equation (11) could be used to rewrite (10) using matrix notation, which will be used in the following. A diagonal matrix $\mathbf{A}$ which describes the texture information will be defined, the size of $\mathbf{A}$ is $(MN) \times (MN)$ and its diagonal elements are

$$A(n,n) = \frac{|C_{x,y} - C_{x-1,y}| + |C_{x,y} - C_{x+1,y}|}{2} \tag{12}$$

where $n = (x-1)M + y$, and $x \in \{1,2,\ldots,M\}, y \in \{1,2,\ldots,N\}$, whereas other elements in the matrix are zero. At this point the distortion E_d is described as

$$\begin{aligned} E_d &\approx \alpha^2 ||\mathbf{A}(\mathbf{D_c} - \mathbf{H}\mathbf{d_c})||^2 \\ &\approx \alpha^2 (\mathbf{D_c}^T - \mathbf{d_c}^T \mathbf{H}^T)\mathbf{A}^T\mathbf{A}(\mathbf{D_c} - \mathbf{H}\mathbf{d_c}). \end{aligned} \tag{13}$$

It is worth noticing that $\mathbf{A}^T\mathbf{A}$ is a diagonal matrix whose entities $A^2(n,n)$ are zeroes when $C_{x,y} = C_{x-1,y} = C_{x+1,y}$ or in other words when texture values around (x,y) is the same in the horizontal direction. The matrix $\mathbf{A}^T\mathbf{A}$ reflects edges contribution in the texture domain along the horizontal direction.

Accordingly, having zero entities on the diagonal of $\mathbf{A}^T\mathbf{A}$, at (n,n), means that the depth distortion $\mathbf{D_c} - \mathbf{H}\mathbf{d_c}$ at position (x,y) is irrelevant to E_d. Thus if the depth map in such position is down-sampled then in the decoder side they can be up-sampled using a simple interpolator without having major impact on E_d, in this case the smaller the baseline the more accurate this conclusion. Since the zero diagonal entities of $\mathbf{A}^T\mathbf{A}$ do not affect E_d, the problem of finding proper down-sampling pattern of the depth map, for the interpolation matrix $\mathbf{H}$, will be reformulated by taking only the non-zero diagonal entities of A into account. In the following the indexes of these entities will be represented by $\mathcal{I} = \{n : A(n,n) \neq 0\}$. Consequently equation (13) will be represented as

$$E_d \approx \alpha^2 \|\overline{\mathbf{A}}(\overline{\mathbf{D}}_\mathbf{c} - \overline{\mathbf{H}}\mathbf{d_c})\|^2 \tag{14}$$

where $\overline{\mathbf{A}}$ is a diagonal matrix whose entities are $A(n,n); n \in \mathcal{I}$. The column vector $\overline{\mathbf{D}}_\mathbf{c}$ is obtained from $\mathbf{D_c}$ and its entities are $D_c(n); n \in \mathcal{I}$; similarly, $\overline{\mathbf{H}}$ is obtained from $\mathbf{H}$ and its entities are $H(n,m); n \in \mathcal{I}, \forall m$. At this stage minimizing the objective function $J = \min_{\mathbf{d_c}} \|\overline{\mathbf{A}}(\overline{\mathbf{D}}_c - \overline{\mathbf{H}}\,\mathbf{d}_c)\|^2$ allows to find the optimal down-sampled depth image for the interpolation matrix $\overline{\mathbf{H}}$. This objective could be achieved by finding the solution to the equation $\frac{\partial J}{\partial \mathbf{d}_c} = 2(\overline{\mathbf{A}}\,\overline{\mathbf{H}})^T(\overline{\mathbf{A}}\,\overline{\mathbf{D}}_c - \overline{\mathbf{A}}\,\overline{\mathbf{H}}\,\mathbf{d}_c) = 0$. Accordingly, the optimal down-sampled depth map can be expressed as

$$\mathbf{d}_c^* = \left(\overline{\mathbf{H}}^T\overline{\mathbf{A}}^2\,\overline{\mathbf{H}}\right)^{-1}\overline{\mathbf{H}}^T\overline{\mathbf{A}}^2\,\overline{\mathbf{D}}_\mathbf{c}. \tag{15}$$

It is worth noticing that $\overline{\mathbf{A}}^2$ is an invertible matrix and that $\overline{\mathbf{H}}^T\,\overline{\mathbf{H}}$ is a positive semidefinite matrix. (15) is the solution for the optimal depth map down-sampling operation for a given interpolation method, where VSD is minimized. It is possible to observe that, the down-sampling approach is derived from the VSD function, which is an approximate function. Nevertheless, this function has been widely accepted because of its good precision.

C. Depth Coding Optimization

In the standardization of 3D-AVC, VSD has been used as an important metric for the selection of the coding mode in depth coding. The VSD evaluation function has been implemented in 3D-AVC and 3D-HEVC. In the case of that depth maps have the same resolution as texture, VSD can be described as (16), shown at the bottom of the page. In 3D-AVC, the default setting is that depth maps are encoded at a reduced resolution, where depth maps have different resolution from texture. The resolution of depth/texture should be adjusted before computing VSD.

In 3D-AVC, VSD can be calculated in both down-sampling domain and up-sampling domain. In down-sampling domain, each depth pixel corresponds to four texture pixels, i.e., (x,y) in depth corresponds to $(2x,2y),(2x,2y+1),(2x+1,2y),(2x+1,2y+1)$ in texture. VSD can be described as (17), shown at the bottom of the page.[3] Here, d is the input down-sampled depth block and $\hat{d}$ represents the decoded down-sampled depth block. In down-sampling domain, only coding distortion is considered when computing VSD. In [24], an approach to compute VSD in up-sampling domain is proposed, which can be described as (18), shown at the bottom of the page. In (18), H denotes the up-sampling operator. In this scheme, each depth block is up-sampled before computing VSD.[4] However, this approach does not consider the down/up-sampling-caused distortion for VSD evaluation (16)-(19) are shown at the bottom of the page.

In this paper, Fig. 2 shows the block diagram of the proposed depth coding scheme. Firstly, the original depth map is down-sampled to low resolution prior to encoding off-line. Then, texture image (full resolution) and depth map (low resolution) will be encoded. In the process of depth coding, the decoded down-sampled depth block is firstly up-sampled using the interpolation method which is also used in the decoder side[5]

[3]The implement can be found in the function *compute_VSD_FlexDepth_UN* of 3D-ATMv9.0.

[4]The implement can be found in the function *compute_VSD16x16, compute_VSD8x8, compute_VSD4x4*.

[5]Any interpolation method could be used given that both encoder and decoder use the same interpolation methods, such as bilinear. In this experiment, we take bilinear sampling method as example.

$$VSD = \sum_{\forall(x,y)} \left[\alpha \underbrace{|D(x,y) - \hat{D}(x,y)|}_{\text{full resolution}} \frac{(|\hat{C}_{x,y} - \hat{C}_{x-1,y}| + |\hat{C}_{x,y} - \hat{C}_{x+1,y}|)}{2} \right]^2 \tag{16}$$

$$VSD = \sum_{\forall(x,y)} \left[\alpha \underbrace{|d(x,y) - \hat{d}(x,y)|}_{\text{low resolution}} \frac{\sum_{k,l=0,1}(|\hat{C}_{2x+k,2y+l} - \hat{C}_{2x-1+k,2y+l}| + |\hat{C}_{2x+k,2y+l} - \hat{C}_{2x+1+k,2y+l}|)}{8} \right]^2 \tag{17}$$

$$VSD = \sum_{\forall(x,y)} \left[\alpha \underbrace{|Hd - H\hat{d}|}_{\text{full resolution}} \frac{|\hat{C}_{x,y} - \hat{C}_{x-1,y}| + |\hat{C}_{x,y} - \hat{C}_{x+1,y}|}{2} \right]^2 \tag{18}$$

$$VSD_{\text{proposed}} = \sum_{\forall(x,y)} \left[\alpha \underbrace{|D - H\hat{d}|}_{\text{full resolution}} \frac{|\hat{C}_{x,y} - \hat{C}_{x-1,y}| + |\hat{C}_{x,y} - \hat{C}_{x+1,y}|}{2} \right]^2 \tag{19}$$

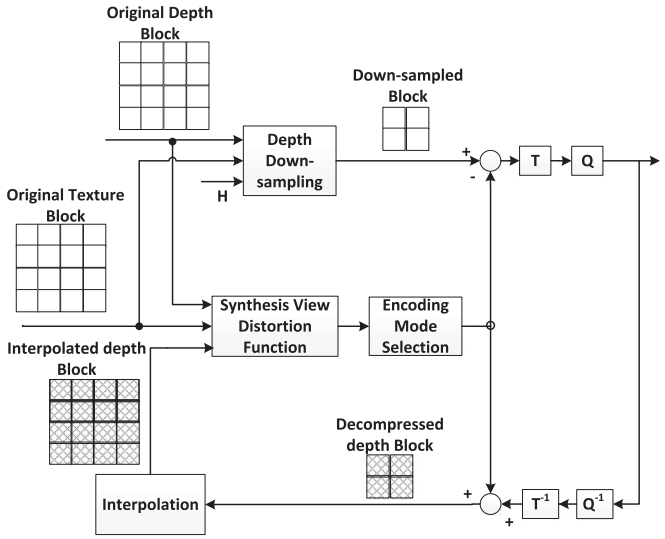

Fig. 2. Diagram of the proposed depth coding scheme.

before computing VSD. Then, the up-sampled depth block as well as the associated decoded texture signal (full resolution) are jointly used to evaluate the VSD. Here, different from the VSD function in up-sampling domain [24], in our proposed scheme, the original full resolution depth map is also used to assist the VSD evaluation. In this case, it takes into account the coding distortion as well as the down/up-sampling-caused distortion. In the proposed coding scheme, the VSD function could be modified as (19), where $\hat{d}$ is the decoded down-sampled depth block, $H\hat{d}$ is the up-sampled version of the decoded depth block and $\hat{C}$ represents the decoded texture block. Different from (18), (19) uses the original depth map D which is full resolution, instead of using the up-sampled depth map Hd. By doing this, the down/up-sampling distortion is also considered in this formula.

III. Experimental Results

This section describes the experimental results of the proposed depth down-sampling and depth coding method. Firstly, the basic experimental setups are depicted. Then, the Rate-Distortion (R-D) performance of the proposed scheme is compared with 3D-AVC. Finally, a complexity comparison against 3D-AVC is concluded to evaluate the proposed scheme.

A. Experimental Setup

In our experiments, the 3DV test sequences used are listed in Table I. The synthesized view is the intermediate view by using View Synthesis Reference Software (VSRS 3.5) [25]. To encode two views of texture and depth with inter-view prediction, 3D-AVC reference software ATM 9.0 [26] is used. The encoder configurations follow the common test conditions (CTC) defined by JCT-3V [27], where the first 100 frames of each sequence including texture and depth views are encoded. Texture views are encoded with full resolution, while depth views are down-sampled prior to encoding by: 1) using the method 1 of JSVM [28] (included in the Starter-kit) which is set as default in 3D-AVC; 2) using non-linear down-sampling [16]; 3) using our proposed down-sampling method. Then, the decoded depth view is up-sampled to the original resolution by using bilinear interpolation method in the decoder side. We tested the proposed method with four quantization parameters (QPs) {26, 31, 36, 41} for the base view, where the same QP is used for both texture and depth views.

TABLE I
3DV Test Sequences Used for the Experiment

Sequences	Encoded View	Synthesized View	Resolution	Tested Frames
Newspaper	2,4	3	1024×768	100
Kendo	1,3	2	1024×768	100
Balloons	1,3	2	1024×768	100
Poznan_Street	3,5	4	1920×1088	100
Poznan_Hall2	3,5	4	1920×1088	100
Undo_Dancer	1,5	3	1920×1088	100
GT_fly	1,5	3	1920×1088	100

TABLE II
Coding Efficiency of Proposed Down-Sampling Stage Versus Anchor, in Terms of Bjontegaard Metrics

Sequences	dBR(%)	dPSNR(dB)
Newspaper	−1.1186	0.0415
Kendo	−0.9003	0.0388
Balloons	−1.9014	0.0665
Poznan_Street	−1.1664	0.0413
Poznan_Hall2	−1.6026	0.056
Undo_Dancer	−1.7769	0.0591
GT_fly	−0.6241	0.0138
Average	−1.2986	0.0453

B. Results

1) Performance of Proposed Down-Sampling Stage: In order to measure the performance of each stage in the proposed scheme, the proposed depth down-sampling method described in Section II-B, is firstly compared with the default depth down-sampling method used in 3D-AVC and depth non-linear down-sampling method, respectively. Based on CTC, texture images are full resolution and depth maps are down-sampled in prior to encoding. Both texture and down-sampled depth are encoded with 3D-AVC. For the "anchor" scheme, depth maps are down-sampled by using the method 1 of JSVM; for the "non-linear" scheme, depth maps are down-sampled by using the nonlinear resampling method [16]; for the "proposed down-sampling", the proposed down-sampling method is applied to sample depth maps.

The Bjontegaard delta bitrate (dBR) and delta Peak Signal-to-Noise Ratio (dPSNR) [29] are used for the objective evaluation of the R-D performance, where the negative values (dBR) indicate reduction of the total bitrate and the positive values (dPSNR) indicate improvement of the synthesized view PSNR. The obtained results of "proposed down-sampling" compared with "anchor" are reported in Table II. The average reduction of total bitrate, which includes both texture and depth, is around 1.3%. As reported in Table III, the average reduction of total

TABLE III
CODING EFFICIENCY OF PROPOSED DOWN-SAMPLING STAGE VERSUS NON-LINEAR, IN TERMS OF BJONTEGAARD METRICS

Sequences	dBR (%)	dPSNR (dB)
Newspaper	−1.3784	0.0416
Kendo	−0.6993	0.0316
Balloons	−0.6597	0.0161
Poznan_Street	−1.1482	0.0327
Poznan_Hall2	−1.3430	0.0247
Undo_Dancer	−0.5289	0.0107
GT_fly	−0.3705	0.0081
Average	−0.8754	0.0236

TABLE IV
CODING EFFICIENCY OF MODIFIED CODING STAGE VERSUS 3D-AVC IN DOWN-SAMPLING DOMAIN, IN TERMS OF BJONTEGAARD METRICS

Sequences	dBR (%)	dPSNR (dB)
Newspaper	−2.9058	0.1012
Kendo	−0.7606	0.0295
Balloons	−0.23	0.0101
Poznan_Street	−0.749	0.0168
Poznan_Hall2	−1.8842	0.0372
Undo_Dancer	−1.6031	0.0618
GT_fly	−0.9681	0.0165
Average	−1.3001	0.039

TABLE V
CODING EFFICIENCY OF MODIFIED CODING STAGE VERSUS 3D-AVC IN UP-SAMPLING DOMAIN, IN TERMS OF BJONTEGAARD METRICS

Sequences	dBR (%)	dPSNR (dB)
Newspaper	−1.9043	0.0651
Kendo	−0.2613	0.0113
Balloons	−0.8784	0.0348
Poznan_Street	−1.1858	0.0176
Poznan_Hall2	−2.8677	0.056
Undo_Dancer	−1.2132	0.0353
GT_fly	−0.4859	0.0037
Average	−1.2567	0.032

TABLE VI
CODING EFFICIENCY OF PROPOSED VERSUS 3D-AVC IN DOWN-SAMPLING DOMAIN, IN TERMS OF BJONTEGAARD METRICS

Sequences	dBR (%)	dPSNR (dB)
Newspaper	−3.1703	0.1032
Kendo	−1.7771	0.0783
Balloons	−4.0126	0.1462
Poznan_Street	−1.7213	0.0436
Poznan_Hall2	−3.1762	0.0934
Undo_Dancer	−2.8926	0.1001
GT_fly	−1.6125	0.0443
Average	−2.6232	0.087

TABLE VII
CODING EFFICIENCY OF PROPOSED VERSUS 3D-AVC IN UP-SAMPLING DOMAIN, IN TERMS OF BJONTEGAARD METRICS

Sequences	dBR (%)	dPSNR (dB)
Newspaper	−2.139	0.0673
Kendo	−1.243	0.0588
Balloons	−4.6656	0.1713
Poznan_Street	−2.4112	0.0486
Poznan_Hall2	−4.1713	0.1107
Undo_Dancer	−2.3508	0.074
GT_fly	−1.1136	0.0311
Average	−2.5849	0.0803

bitrate is 0.88% compared the "proposed down-sampling" with "non-linear down-sampling". As can be observed from Tables II and III, for the sequences "*Newspaper*" and "*Poznan_Hall2*", the R-D improvement is larger than that of other sequences. Because there is more boundary information in these sequences, which plays a significant role during virtual view synthesis. By taking into account VSD during down-sampling process, more boundary information can be preserved. Consequently, higher synthesized view quality can be obtained.

2) Performance of Proposed Coding Stage: In Section II-C, a modified VSD function is proposed, where the reconstructed depth block is up-sampled and used to evaluate VSD before coding mode decision. Thus, the down/up-sampling distortion and coding distortion can be jointly considered to select the best coding mode. In this section, the performance of the proposed coding scheme (denoted as "modified coding stage") is compared with 3D-AVC. The original texture video and the default down-sampled depth maps are used as input in both "3D-AVC" and "modified coding stage". As described in II-C, the distortion function based on the up-sampling domain in [24] has been proposed and integrated into 3D-AVC. Hence, the proposed coding scheme is used to compare with 3D-AVC in the down-sampling domain (17) and up-sampling domain (18) in our experiment, respectively. The average of bitrate reduction against "3D-AVC with down-sampling domain" is around 1.3% as reported in Table IV, while the average of bitrate reduction against "3D-AVC with up-sampling domain" is near to 1.26% as shown in Table V. In general, the down-sampling operation introduces more distortion into the boundary regions than the homogeneous regions. By up-sampling the decoded depth block before evaluating the synthesized view distortion, the distortion introduced by down/up-sampling can also be taken into account in the coding selection step. Thereby, for the sequence "*Newspaper*" and "*Poznan_Hall2*" with much boundary information, larger gains can be obtained compared with other sequences. Nevertheless, for the other sequences, a small gain can also be seen, such as "*Kendo*" and "*GT_fly*". It means that down/up-sampling distortion plays a role in influencing the R-D performance of resolution-reduction depth coding.

3) Overall Performance: In the proposed scheme, through minimizing the down-sampling-caused and coding-caused VSD, the best R-D performance can be achieved. The results against 3D-AVC are shown in Tables VI and VII, where "proposed" indicates that both the proposed down-sampling and proposed coding schemes are jointly applied. As can be observed from Table VI, the average bitrate reduction against "3D-AVC with down-sampling domain" is 2.62%, while the average bitrate

TABLE VIII
COMPLEXITY COMPARISON AGAINST ANCHOR

Sequence	Original coding stage (Enc. Time)	Modified coding stage (Enc. Time)	Proposed (Enc. Time)
Newspaper	105.99%	122.44%	117.08%
Kendo	88.80%	92.26%	99.44%
Balloons	96.43%	110.14%	107.52%
Poznan_Street	73.23%	122.36%	123.05%
Poznan_Hall2	58.98%	80.55%	86.74%
Undo_Dancer	94.34%	109.94%	116%
GT_fly	128.24%	141.30%	126.89%
Average	92.29%	111.28%	110.96%

reduction is near to 2.58% against "3D-AVC with up-sampling domain" in Table VII. As expected in this experiment, we found that by jointly using the proposed depth down-sampling and coding scheme, the sequences with much boundary information could achieve larger R-D performance gain, compared with separately using one of two proposed approaches; while for the sequences with little boundary information, the improvement is limited. But against anchor, the gains are obvious. We can conclude that by considering the down/up-sampling-caused and coding-caused VSD in the whole process of down-sampling and coding, a better compression for depth maps and higher objective quality for synthesized views can be achieved.

C. Complexity of Coding

In this section, the coding complexity of the proposed scheme is compared in terms of running time. Here, we set the coding configuration with CTC as anchor, where the default down-sampled depth maps are encoded with 3D-AVC. The testing environment is an Intel Core2 3.0 GHz with 4 GB RAM. In this experiment, three coding configurations are considered:

1) the down-sampled depth maps using the proposed depth down-sampling method are encoded with 3D-AVC;
2) the default down-sampled depth maps are encoded with the proposed coding scheme;
3) the down-sampled depth maps using the proposed depth down-sampling method are encoded with the proposed coding scheme.

The running time results of the proposed coding scheme is listed in Table VIII, where the "original coding stage" presents the down-sampled depth maps using the proposed depth-sampling method are encoded with 3D-AVC, "modified coding stage" is the default down-sampled depth maps are encoded with the proposed coding scheme, and "proposed" denotes that the down-sampled depth maps using the proposed depth down-sampling method are encoded with the proposed coding scheme. In average, the complexity of "original coding stage" is relatively similar with that of anchor; whereas for "modified coding stage" and "proposed", the run time doesn't largely increase. In some cases, the run time of encoding is less than the anchor. Since the mode decision is based on the modified mode selection function, more Skip and Direct modes might be selected during encoding; Therefore, the proposed scheme strikes a better trade-off between the coding gain and complexity.

IV. CONCLUSION

In this paper, we have proposed an optimal depth map down-sampling and coding scheme based on synthesized view distortion. Different from the conversional down-sampling methods, we introduced the synthesized view distortion into the depth map down-sampling and coding procedure. Consequently, we obtained a better R-D performance by minimizing the down-sampling and synthesized view distortion. The effectiveness of the proposed approach was demonstrated by the results.

REFERENCES

[1] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," *Proc. SPIE*, vol. 5291, pp. 93–104, May 2004.

[2] A. Smolic and D. McCutchen, "3DAV exploration of video-based rendering technology in MPEG," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 348–356, Mar. 2004.

[3] J. Fu *et al.*, "Kinect-like depth data compression," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1340–1352, Oct. 2013.

[4] A. Tikanmaki, A. Gotchev, A. Smolic, and K. Miller, "Quality assessment of 3D video in rate allocation experiments," in *Proc. IEEE Int. Symp. Consum. Electron.*, Apr. 2008, pp. 1–4.

[5] S. Ma, S. Wang, and W. Gao, "Low complexity adaptive view synthesis optimization in HEVC based 3D video coding," *IEEE Trans. Multimedia*, vol. 16, no. 1, pp. 266–271, Jan. 2014.

[6] F. Shao, G. Jiang, M. Yu, K. Chen, and Y.-S. Ho, "Asymmetric coding of multi-view video plus depth based 3D video for view rendering," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 157–167, Feb. 2012.

[7] K. Klimaszewski, K. Wegner, and M. Domanski, *Influence of Views and Depth Compression Onto Quality of Synthesized Views*, ISO/IEC JTC1/SC29/WG11 MPEG2009, doc. no. M16758, 2009.

[8] *Test Model for AVC Based 3D Video Coding*, ISO/IEC JTC1/SC29/WG11 MPEG2012/N12558, Feb. 2012.

[9] E. Ekmekcioglu, S. T. Worrall, and A. M. Kondoz, "Bit-rate adaptive downsampling for the coding of multi-view video with depth information," in *Proc. 3DTV Conf., True Vis.-Capture, Transmiss. Display 3D Video*, May 2008, pp. 137–140.

[10] K.-J. Oh, S. Yea, A. Vetro, and Y.-S. Ho, "Depth reconstruction filter and down/up sampling for depth coding in 3D video," *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 747–750, 2009.

[11] S. Liu, P. Lai, D. Tian, C. Gomila, and C. W. Chen, "Joint trilateral filtering for depth map compression," *Proc. SPIE*, vol. 7744, 2010, Art. no. 77440F.

[12] H. Deng, L. Yu, and Z. Xiong, "Edge-preserving interpolation for down/up sampling-based depth compression," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep.–Oct. 2012, pp. 1301–1304.

[13] M. Wildeboer, T. Yendo, M. P. Tehrani, T. Fujii, and M. Tanimoto, "Color based depth up-sampling for depth compression," in *Proc. Picture Coding Symp.*, 2010, pp. 170–173.

[14] Q. Liu, Y. Yang, R. Ji, Y. Gao, and L. Yu, "Cross-view down/up-sampling method for multiview depth video coding," *IEEE Signal Process. Lett.*, vol. 19, no. 5, pp. 295–298, 2012.

[15] P. Aflaki, D. Rusanovskyy, and M. M. Hannuksela, *3DV-CE3: Non-linear Depth Map Downsampling for 3DV-ATM Coding (Pre-processing)*, ITU-T SG 16 WP 3 ISO/IEC JTC 1/SC 29/WG 11 MPEG2012/M26072, Jul. 2012.

[16] P. Aflaki, M. Hannuksela, D. Rusanovskyy, and M. Gabbouj, "Nonlinear depth map resampling for depth-enhanced 3D video coding," *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 87–90, Jan. 2013.

[17] Y. Zhang, D. Zhao, J. Zhang, R. Xiong, and W. Gao, "Interpolation-dependent image downsampling," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3291–3296, Nov. 2011.

[18] W.-S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, "Depth map distortion analysis for view rendering and depth coding," in *Proc. 16th IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 721–724.

[19] B. T. Oh and K.-J. Oh, "View synthesis distortion estimation for AVC-and HEVC-compatible 3D video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 6, pp. 1006–1015, Jun. 2014.

[20] B. T. Oh, J. Lee, and D.-s. Park, "Depth map coding based on synthesized view distortion function," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 7, pp. 1344–1352, Nov. 2011.
[21] P. Gao and W. Xiang, "Rate-distortion optimized mode switching for error-resilient multi-view video plus depth based 3D video coding," *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 1797–1808, Nov. 2014.
[22] B. Macchiavello, C. Dorea, E. Hung, G. Cheung, and W.-T. Tan, "Loss-resilient coding of texture and depth for free-viewpoint video conferencing," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 711–725, Apr. 2014.
[23] A. M. Bruckstein, M. Elad, and R. Kimmel, "Down-scaling for better transform compression," *IEEE Trans. Image Process.*, vol. 12, no. 9, pp. 1132–1144, Sep. 2003.
[24] J. L. Byung Tae Oh and D. S. Park, "3D-CE8.a results on view synthesis optimization using distortion in synthesized views by Samsung," presented at the ISO/IEC JTC1/SC29/WG11 MPEG2012/M24826, Geneva, Switzerland, May 2012.
[25] M. Tanimoto, T. Fujii, and K. Suzuki, *View Synthesis Algorithm in View Synthesis Reference Software 3.5 (VSRS3. 5)*, ISO/IEC JTC1/SC29/WG11 (MPEG), doc. no. M16090, 2009.
[26] "3D-AVC Reference Software ATM 9.0," [Online]. Available: http://mpeg3dv.research.nokia.com/svn/mpeg3dv/tags/3DV-ATMv9.0/
[27] D. Rusanovskyy, K. Mller, and A. Vetro, "Common test conditions of 3DV core experiments," presented at the 5th Meeting Joint Collaborative Team Video Coding ITU-T SG 16 WP 3 and ISO/IEC JTC1/SC29/WG11, Vienna, Austria, Jul. 27–Aug. 2, 2013.
[28] J. Reichel, H. Schwarz, and M. Wien, *Joint Scalable Video Model 11 (JSVM 11)*, doc. no. JVT-X202, Joint Video Team, 2007.
[29] G. Bjontegaard, "Calculation of average PSNR differences between R-D curves," presented at the 13th Meeting ITU-T SC16/SG16 VCEG, Austin, TX, USA, Apr. 2001.

Chao Yao received the B.S. degree in computer science from Beijing Jiaotong University (BJTU), Beijing, China, in 2009, and the Ph.D. degree from the Institute of Information Science, BJTU, in 2016.

From 2014 to 2015, he served as a Visiting Ph.D. student at Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland. Since July 2016, he has been with the Institute of Sensing Technology and Business, Beijing University of Posts and Telecommunications, Beijing, China. His current research interests include image and video processing, computer vision, and robot technique.

Jimin Xiao received the B.S. and M.E. degrees in telecommunication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2004 and 2007, respectively, and the Ph.D. degree in electrical engineering and electronics from the University of Liverpool, Liverpool, U.K., in 2013.

From November 2013 to November 2014, he was a Senior Researcher with the Department of Signal Processing, Tampere University of Technology, Tampere, Finland, and an External Researcher with the Nokia Research Center, Tampere, Finland. Since December 2014, he has been a Faculty Member with Xi'an Jiaotong-Liverpool University, Suzhou, China. His research interests include image and video processing, computer vision, and deep learning.

Tammam Tillo (S'03–M'05–SM'12) was born in Damascus, Syria. He received the Engineer Diploma in electrical engineering from the University of Damascus, Damascus, Syria, in 1994, and the Ph.D. degree in electronics and communication engineering from Politecnico di Torino, Turin, Italy, in 2005.

In 2004, he was as a Visiting Researcher with the Ecole Polytechnique Federale de Lausanne, Switzerland, and from 2005 to 2008, he was a Postdoctoral Researcher with the Image Processing Lab, Politecnico di Torino. For few months, he was an Invited Research Professor with the Digital Media Lab, Sungkyunkwan University, Seoul, South Korea. He joined Xi'an Jiaotong-Liverpool University (XJTLU), Suzhou, China, in 2008. From 2010 to 2013, he was the Head of the Department of Electrical and Electronic Engineering, XJTLU University, and he was the Acting Head of the Department of Computer Science and Software Engineering from 2012 to 2013. He currently serves as an Expert Evaluator for several national-level research programs. His research interests include robust transmission of multimedia data, image and video compression, and hyperspectral image compression.

Yao Zhao (M'06–SM'12) received the B.S. degree in radio engineering from Fuzhou University, Fuzhou, China, in 1989, the M.E. degree in radio engineering from Southeast University, Nanjing, China, in 1992, and the Ph.D. degree from the Institute of Information Science, Beijing Jiaotong University (BJTU), China, in 1996.

He became an Associate Professor with BJTU in 1998 and a Professor in 2001. From 2001 to 2002, he was a Senior Research Fellow with the Information and Communication Theory Group, Faculty of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands. In October 2015, he visited Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland. He is currently the Director of the Institute of Information Science, BJTU. His current research interests include image/video coding, digital watermarking and forensics, and video analysis and understanding.

Dr. Zhao is a Fellow of the IET. He serves on the Editorial Boards of several international journals, including as an Associate Editor of the IEEE TRANSACTIONS ON CYBERNETICS and the IEEE SIGNAL PROCESSING LETTERS, and an Area Editor of *Signal Processing: Image Communication*. He was named a Distinguished Young Scholar by the National Science Foundation of China in 2010, and was elected as a Chang Jiang Scholar of the Ministry of Education of China in 2013.

Chunyu Lin was born in Liaoning Province, China. He received the Ph.D. degree from Beijing Jiaotong University, Beijing, China, in 2011.

From 2009 to 2010, he was a Visiting Researcher with the ICT Group, Delft University of Technology, Delft, The Netherlands. From 2011 to 2012, he was a Postdoctoral Researcher with the Multimedia Laboratory, Ghent University, Ghent, Belgium. He is currently an Associate Professor with Beijing Jiaotong University. His current research interests include image/video compression and robust transmission, 2D-to-3D conversion, and 3D video processing.

Huihui Bai received the Ph.D. degree in signal and information processing from Beijing Jiaotong University (BJTU), Beijing, China, in 2008.

She is currently an Associate Professor with the Institute of Information Science, BJTU. She is leading or participating in several research projects such as the 973 Program, the 863 Program, the National Natural Science Foundation of China, the Beijing Natural Science Foundation, and the Jiangsu Provincial Natural Science Foundation. Her current research interests include video coding technologies and standards such as HEVC, 3D video compression, multiple description video coding, and distributed video coding.